This is the submitted version of this article. Published as:

Barnett, Adrian G. and Koper, Nicola and Dobson, Annette J. and Schmiegelow, Fiona and Manseau, Micheline (2010) *Using information criteria to select the correct variance–covariance structure for longitudinal data in ecology*. Methods in Ecology & Evolution.

Using information criteria to select the correct variance–covariance structure for longitudinal data in ecology

Running title: Selecting the correct variance–covariance

Adrian G. Barnett
Institute of Health and Biomedical Innovation
& School of Public Health
Queensland University of Technology, 60 Musk Avenue
Kelvin Grove, Queensland, 4059, Australia
Tel.: +61-7-3138 6010, Fax: +61-7-3138 6030
E-mail: a.barnett@qut.edu.au
Nicola Koper
Natural Resources Institute, University of Manitoba
70 Dysart Road, Winnipeg, Manitoba, Canada, R3T 2N2
Annette J. Dobson
School of Population Health, University of Queensland
Herston, QLD 4006, Australia
Fiona Schmiegelow
Department of Renewable Resources, University of Alberta
Edmonton, Alberta, Canada
Micheline Manseau
Natural Resources Institute, University of Manitoba
70 Dysart Road, Winnipeg, Manitoba, Canada, R3T 2N2

Word count: 6998

Abstract

1. Ecological data sets often use clustered measurements or use repeated sampling in a longitudinal design. Choosing the correct covariance structure is an important step in the analysis of such data, as the covariance describes the degree of similarity among the repeated observations.

2. Three methods for choosing the covariance are: the Akaike information criterion (AIC), the quasi-information criterion (QIC), and the deviance information criterion (DIC). We compared the methods using a simulation study and using a data set that explored effects of forest fragmentation on avian species richness over 15 years.

3. The overall success was 80.6% for the AIC, 29.4% for the QIC and 81.6% for the DIC. For the forest fragmentation study the AIC and DIC selected the unstructured covariance, whereas the QIC selected the simpler autoregressive covariance. Graphical diagnostics suggested that the unstructured covariance was probably correct.

4. We recommend using DIC for selecting the correct covariance structure.
Key-words: Bayesian methods, correlated data, covariance structure, information criteria, generalized estimating equation, longitudinal data

Introduction

Ecological data are often clustered or otherwise correlated, either because of intrinsic ecological patterns or because of the way data were collected. This can occur by clustering sub-samples within study sites (Koper & Schmiegelow, 2006), by repeatedly sampling individuals or sites (longitudinal studies, e.g. Reynolds (2004)), or because of phylogenetic relationships among focal species (Duncan, 2004). Such clustering should not be seen as a flaw in the study design, as the repeated nature of the data means that such studies are strongly placed to examine ecological changes over time (e.g., Schmiegelow et al (1997)). Also, clustered sampling designs are often intrinsic to the nature of the ecological system. For example, the need for a nested sampling design to explore effects of habitat structure at multiple spatial scales has long been recognized in landscape ecology (Wiens, 1989).

Correlation among clustered or repeated measurement data means that independence can no longer be assumed among all observations. Hence, most standard statistical analyses cannot be used to analyze this type of data. If standard analyses are used, the likelihood of Type I errors is increased (Clifford et al, 1989). A number of approaches are available for analyzing correlated data, however, and their use is becoming increasingly common in ecology. Mixed models (e.g., Krawchuk & Taylor (2003); Gillies et al (2006)), generalized linear models with generalized estimating equations (e.g., Dreitz et al (2004); Driscoll et al (2005)) and Bayesian models (e.g., Schneider et al (2006); Helser et al (2007)) have all been applied to ecological data to control for clustering or repeated measures. However, selecting which approach is optimal for analysis of a particular study is not trivial, because each of these methods has a different conceptual paradigm, and its own strengths and weaknesses.

A key step in the analysis of correlated data is to determine the appropriate covariance structure, which describes the form (or structure) of the correlation among data points within clusters (Fitzmaurice et al, 2004). This is important because the overall model fit, the parameter estimates, and their standard errors can be sensitive to the model covariance structure (Fitzmaurice et al, 2004). The covariance is often given a simplifying structure, as this reduces the number of parameters and can improve model convergence.

A number of different covariance structures are available that cover a range of assumptions about the associations between responses from the same cluster. An independent covariance would be appropriate when none of responses are correlated. An exchangeable covariance would be appropriate when responses from the same cluster are equally correlated, regardless of the distance between responses. An autoregressive covariance would be appropriate when the correlation between responses decays with distance. An unstructured covariance would be appropriate when the correlation between responses is comparatively complex, or when the variance is heterogeneous (Grady, 1995).

The criteria used to identify which covariance structure gives the best trade-off between model fit and complexity differ between maximum-likelihood mixed effects models, generalized estimating equations, and mixed effects models fitted using a Bayesian paradigm. Our objective was to compare the performance of alternative information criteria for selecting among alternative covariance structures.

We compared three criteria for finding the optimal covariance: the Akaike information criterion (AIC, using mixed models) the quasi-information criterion (QIC, using generalized estimating equations), and the deviance information criterion (DIC, using Bayesian models). AIC has been used extensively for model selection in ecological research, while the use of QIC and DIC seem to be gradually increasing. Our objective was to determine the optimal criterion under a range of conditions typical of ecological data. We first used a simulation study, using data with known covariance structures, to compare the performance of the information criteria in selecting the correct covariance. We then compared the criteria using an empirical data set describing effects of time since forest fragmentation on avian richness.

Materials and methods

We start with some notation and assumptions. We label the repeated data from cluster i using $\mathbf{Y}_i = Y_{i1}, Y_{i2} \ldots, Y_{im}$, so there are m responses per cluster, and we label the total number of clusters as N. For simplicity we only consider Normally distributed response data (i.e., $\mathbf{Y}$ has a multivariate Normal distribution), and balanced data so each cluster has the same number of responses m. We assume that the repeated data were generated by sampling the same location (or measuring the same subject) at multiple times ($t = 1, \ldots, m$). However, the methods could be applied to non-longitudinal data, such as responses from the same family (e.g., siblings), or samples that are spatially clustered.

Variance–covariance matrices

We define the variance–covariance of the responses in a cluster, $\text{Var}(\mathbf{Y}_i)$, using the $m \times m$ symmetric matrix

$$\mathbf{V}_i = \begin{bmatrix} \sigma_{i1}^2 & \sigma_{i1}\sigma_{i2} & \ldots & \sigma_{i1}\sigma_{im} \\ \sigma_{i2}\sigma_{i1} & \sigma_{i2}^2 & & \sigma_{i2}\sigma_{im} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{im}\sigma_{i1} & \sigma_{im}\sigma_{i2} & \ldots & \sigma_{im}^2 \end{bmatrix} \tag{1}$$

The diagonal elements of $\mathbf{V}_i$ are variances and the off-diagonal elements are covariances. Equation (1) involves $m(m+1)/2$ covariance parameters per cluster for $\mathbf{V}_i$. To reduce the total number of parameters it is common to assume that: i) each cluster has the same variance–covariance matrix, and ii) that the matrix has some structure.

There are a large number of covariance structures to choose from. In this paper we focus on the following four: independent, exchangeable, autoregressive and unstructured. These four structures cover a range of different scenarios for the pattern of covariance, and are those most commonly available in statistics packages. For example, we might assume that the covariance between all observations from the same cluster is constant, and that the variance remains constant over time. The variance–covariance matrix would then be:

$$\mathbf{V}_i = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \dots & \sigma^2\rho \\ \sigma^2\rho & \sigma^2 & & \sigma^2\rho \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2\rho & \sigma^2\rho & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix} \tag{2}$$

where $-1 < \rho < 1$ measures the constant within-cluster correlation, and $\sigma^2 > 0$ the variance. This structure has only two covariance parameters (σ^2, ρ) and is known as the exchangeable covariance matrix because the observations from any cluster could be re-arranged (exchanged) in time, and the covariance between observations would remain the same. The right hand side of equation (2) has split the variance–covariance matrix into a variance parameter and correlation matrix.

The autoregressive structure assumes a steady decay in correlation with increasing time or distance between observations. It is common to use an autoregressive model of order one, labeled AR(1), which has one correlation parameter and one variance (as does the exchangeable covariance). The correlation between observations from the same cluster at times r and s is

$\rho^{|r-s|}$ as $|\rho| < 1$. So, the correlation decreases as the distance $|r-s|$ between times increases.

The unstructured covariance assumes that no two pairs of observations are equally correlated, and that there is no "structure" between neighboring values in the matrix. Additionally, it also allows different variance terms along the diagonal of the matrix. Notationally, it is the matrix in equation (1) without the index i. The number of parameters is $m(m+1)/2$, where m is the number of responses within the cluster, so the number of parameters can be large for this covariance matrix.

At the opposite end of the spectrum from the unstructured covariance is the independent covariance, which assumes no correlation between observations. This is equivalent to the exchangeable covariance (2) with $\rho = 0$. This structure is useful for determining whether more complex structures improve model fit.

Mixed effects models

Mixed effects models are a popular method for analysing correlated data. They are called "mixed" models as they are a mix of fixed and random effects. As an example, a linear regression model with a single time-dependent covariate X_{it} is

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_i + \varepsilon_{it}, \qquad i = 1, \ldots, N, \ t = 1, \ldots, m.$$

In this model β_0 and β_1 are fixed parameters as they are the same for all clusters, whereas γ_i is a random parameter as it varies by cluster. In this case γ_i is a random intercept, which is a useful way of modelling the similarity in responses from the same cluster. It is also possible to model the similarity in responses using the error terms ($\boldsymbol{\varepsilon}$) if we define them using a multivariate Normal distribution

$$\boldsymbol{\varepsilon}_i \sim \text{MVN}(\mathbf{0}, \mathbf{V}), \qquad i = 1, \ldots, N,$$

where $\mathbf{V}$ is the variance–covariance matrix (which is the same for all clusters). A model without any random effects but with a variance–covariance matrix is called a "covariance pattern model" by (Fitzmaurice et al, 2004, Chapter 7), and these are the mixed models that we use here.

Generalized Estimating Equations (GEEs)

Generalized estimating equations (GEEs) can be used to model correlated data with the variance–covariance matrix $\mathbf{V}$ by iteratively solving the score equation:

$$\sum_{i=1}^{N} \left(\frac{d\boldsymbol{\mu}_i(\boldsymbol{\beta})}{d\boldsymbol{\beta}} \right) \mathbf{V}_i^{-1} \left(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}) \right) = 0, \tag{3}$$

where $\boldsymbol{\mu}_i(\boldsymbol{\beta})$ is the fitted mean, which is given by $g\left(\mu_{it}(\boldsymbol{\beta})\right) = x_{it}\boldsymbol{\beta}$ for covariates $\mathbf{x} = \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{im}$ and regression parameters $\boldsymbol{\beta} = \beta_1, \ldots, \beta_p$.

GEEs are fitted using a quasi-likelihood method rather than the maximum likelihood (Hardin & Hilbe, 2003, page 34). The estimates from a GEE analysis are robust to mis-specification of the covariance matrix (Liang & Zeger, 1986), so even when an independent covariance matrix is used the regression parameter estimates are consistent. Using a hypothesised covariance matrix (sometimes called the "working" covariance) that is closer to the true covariance improves the precision of the estimates (i.e., reduces standard errors) (Diggle et al, 2002; Fitzmaurice et al, 2004). Using an incorrect working covariance can lead to failed convergence or biased standard errors (Hilbe, 2009).

Bayesian methods for correlated data

We can use Bayesian methods to estimate the regression parameters and variance–covariance structure. An advantage of a Bayesian model is the use of Markov chain Monte Carlo (MCMC) estimation for the regression and variance–covariance parameters. This results in more easily interpretable statistical findings than traditional analytical methods (Dobson & Barnett, 2008, Chapter 12).

One of the main differences between classical statistical methods and Bayesian methods is the use of a prior distribution (Dobson & Barnett, 2008, Chapter 12). Priors can be used to model existing knowledge (e.g., a positive correlation between species richness and island size), or to incorporate information about the model or study design.

For the Bayesian approach, the variance–covariance structure can be parameterized in terms of the inverse of the variance–covariance matrix (Spiegelhalter et al, 2007). An unstructured covariance can be modeled by using a Wishart prior

$$\mathbf{V}^{-1} \sim \mathrm{W}(\mathbf{\Sigma}, \nu),$$

where $\mathbf{\Sigma}$ is the prior estimate of the variance–covariance matrix and ν is the degrees of freedom, which controls the weight given to the prior. The inverse Wishart is the conjugate prior for the multivariate Normal distribution, and gives covariance matrices that are symmetric and positive definite.

An autoregressive variance–covariance matrix can be formulated by taking advantage of the structure of the inverse matrix. The term for row r and column s of the inverse covariance matrix is,

$$V_{rs}^{-1} = \begin{cases} \tau, & r = s = 1, m \\ \tau(1+\rho^2), & r = s = 2, \dots, m-1 \\ -\tau\rho, & r = 1, \dots, m-1,\ s = r+1 \\ -\tau\rho, & s = 1, \dots, m-1,\ r = s+1 \\ 0, & \text{otherwise} \end{cases}$$

This structure has two unknown parameters, τ and ρ.

The exchangeable variance–covariance matrix can be formulated using the inverse of the matrix in equation (2),

$$V_{rs}^{-1} = \begin{cases} [1+(m-2)\rho]/\gamma, & r = s = 1, \dots, m \\ -\rho/\gamma, & r, s = 1, \dots, m,\ r \neq s \end{cases}$$

where $\gamma = \sigma^2[1+(m-2)\rho+(m-1)\rho^2]$. This structure also has two unknown parameters, τ and ρ.

The independent variance–covariance matrix has the simple form,

$$V_{rs}^{-1} = \begin{cases} 1/\sigma^2, & r = s = 1, \dots, m \\ 0, & \text{otherwise} \end{cases}$$

Akaike information criterion

A commonly used statistic with models derived using maximum likelihood is the Akaike information criterion (AIC, Akaike (1974)). The equation for the fixed effects AIC is

$$\mathrm{AIC} = -2\log L + 2p_A, \tag{4}$$

where L is the likelihood and p_A the total number of parameters. The AIC is a trade-off between a good fit to the model (measured by the likelihood), and a penalty for complexity (calculated using the number of parameters). We can calculate the AIC for different models describing the same data, and the one with the lowest AIC is interpreted as the best model.

Quasi-information criterion

Although the AIC can be used in association with mixed models, it cannot be used with GEEs to select either the optimal set of explanatory variables or covariance matrix, because GEE estimation is based on the quasi-likelihood rather than the maximum likelihood. The quasi-likelihood counterpart to the AIC is the QIC, or the "quasi-likelihood under the independence model information criterion" (Pan, 2001). The QIC was derived from the AIC and is conceptually similar. An equation for the QIC is:

$$\mathrm{QIC}_P = -2Q(\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}, \mathbf{I}) + 2 \times \mathrm{trace}\left[\left(\hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}, \mathbf{I})\right)^{-1} \hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}, \hat{\mathbf{V}})\right], \qquad (5)$$

where $Q(\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}, \mathbf{I})$ is the quasi-likelihood calculated using an independent covariance $\mathbf{I}$, but with the regression parameter estimates ($\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}$) fitted using the estimate of the hypothesized covariance matrix $\hat{\mathbf{V}}$ (Hardin & Hilbe, 2003, page 140). Like the AIC, the QIC is a trade-off between a good fit to the model, as measured by the quasi-likelihood, and a penalty for over-complexity as measured by the trace. The optimal variance–covariance matrix is that which gives the smallest QIC.

The terms $\hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}, \mathbf{I})$ and $\hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}})$ are $p \times p$ matrices, where p is the number of regression parameters. $\hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}, \mathbf{I})$ is the *model-based* covariance matrix for the estimated regression parameters using an independent covariance matrix. The general formula for the model-based covariance is

$$\hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}}) = \left[\sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}}\right)^T \hat{\mathbf{V}}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}}\right]^{-1}.$$

Thus $\hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}})$ is covariance matrix for the regression parameters using the hypothesized covariance matrix. The other term, $\hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}})$ is also known as the robust or sandwich estimate (Dobson & Barnett, 2008), because it formed as a "sandwich" by the model-based estimate:

$$\hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}}) = \hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}, \mathbf{V}) \mathbf{C} \hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}, \mathbf{V}), \qquad (6)$$

$$\text{where } \mathbf{C} \;=\; \sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}}\right)^T \hat{\mathbf{V}}_i^{-1} \left(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i(\boldsymbol{\beta})\right) \left(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i(\boldsymbol{\beta})\right)^T \hat{\mathbf{V}}_i^{-1} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}}\right).$$

The estimates of $\hat{\boldsymbol{\beta}}$ using $\hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}})$ are robust to the mis-specification of $\mathbf{V}$, whereas those using the model-based covariance are not.

A slightly different version of the QIC suggested by Hardin & Hilbe (2003) is

$$\mathrm{QIC}_{HH} = -2Q(\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}, \mathbf{I}) + 2 \times \mathrm{trace}\left[\left(\hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}_{\mathbf{I}}, \mathbf{I})\right)^{-1} \hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}, \hat{\mathbf{V}})\right], \qquad (7)$$

so the first term in the trace differs from Equation (5). Following Hin & Wang (2009) we have labeled Equation (5) as the QIC_P as it follows Pan's original formulation, and Equation (7) as the QIC_{HH} as it was designed by Hardin and Hilbe. Hin & Wang (2009) stated that the difference in the $\mathrm{QIC}_P(R)$ and $\mathrm{QIC}_{HH}(R)$ is only $O(m^{-1/2})$, which is small for data sets with even only a moderate number of clusters ($m \geq 50$). The QIC_P and QIC_{HH} are identical for the independent matrix.

If the covariate matrix $\mathbf{x}$ does not contain at least one covariate that is both: i) time-dependent (Diggle et al, 2002, Chapter 12), and ii) cluster-specific, then the sandwich estimate $\hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}, \mathbf{I})$ using an independent covariance is identical to the estimate using an exchangeable covariance, $\hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}, \mathbf{V})$. This is because cancelation of the terms involving $\partial\hat{\boldsymbol{\mu}}_i/\partial\hat{\boldsymbol{\beta}}$ in equation (6) leads to both covariance structures leading to the same regression parameter estimates. This means that values of the QIC_P and

QIC_{HH} will be the same for an independent covariance structure and an exchangeable one. This is an obvious drawback, as neither of the QIC statistics can distinguish between these two structures, which have very different interpretations.

Deviance information criterion

The deviance information criterion (DIC) is a generalisation of the AIC for Bayesian analysis (Spiegelhalter et al, 2002). The formula for the DIC is similar to the formula for the AIC (4)

$$\mathrm{DIC} = D(\mathbf{Y}|\overline{\boldsymbol{\beta}}) + 2p_D, \tag{8}$$

where $D(\mathbf{Y}|\overline{\boldsymbol{\beta}})$ is the deviance using the estimates of the regression parameters means averaged over the MCMC samples ($\overline{\boldsymbol{\beta}}$). The effective number of parameters is p_D and is not necessarily an integer; it can be thought of as the amount of information needed to fit the model. It is estimated using

$$p_D = \overline{D(\mathbf{Y}|\boldsymbol{\beta})} - D(\mathbf{Y}|\overline{\boldsymbol{\beta}}).$$

where $\overline{D(\mathbf{Y}|\boldsymbol{\beta})}$ is the average deviance over all values of $\boldsymbol{\beta}$. The effective number of parameters is thus the mean deviance minus the deviance at the means.

Similarly to the AIC and QIC, the DIC aims to be a trade-off between a good fit to the model (as measured by the deviance), and a penalty for complexity measured by the effective number of parameters.

Comparisons of AIC, QIC and DIC

The three information criteria, Equations (4), (5) and (8), have an identical form, and also share the same goal: to balance model fit and complexity.

The effective number of parameters is estimated for the QIC and DIC, whereas for the AIC it is fixed as it is based on the actual number of parameters. When using the unstructured covariance the number of estimated parameters for the covariance is $m(m+1)/2$. However, some of these parameters may be correlated, which would reduce the complexity. This reduction in complexity can potentially be captured by the QIC and DIC, but not by the AIC equation used here (4).

Data

We compared the performance of the three information criteria using data from a simulation study (with known covariance structure), and empirical data from an ecological study. In this section we describe these two data sources.

Simulation study data

The simulated data used 30 clusters, 8 responses per cluster with no missing data, and a single regression parameter β. We simulated data using the following multivariate Normal distribution and regression equation

$$\begin{aligned} \mathbf{Y}_i &\sim \mathrm{MVN}(\boldsymbol{\mu}_i, \mathbf{V}), \qquad i = 1, \ldots, 30, \\ \mu_{it} &= \beta X_{it}, \qquad t = 1, \ldots, 8. \end{aligned} \tag{9}$$

We used four different covariance structures for $\mathbf{V}$: independent, exchangeable, autoregressive and unstructured. For each covariance structure we ran two regression models (9). One regression model used a fixed covariate common to all clusters, $X_{it} = t$. The other regression model used a random covariate, $X_{it} \sim \mathrm{N}(0, 1)$, which was both cluster-specific and time-dependent. For both regression models we used $\beta = 0.3$.

For each combination of covariance type and regression model we ran 100 simulations. For the exchangeable data we used two different values for the within-cluster correlation: a moderate correlation of $\rho = 0.5$ and a weak correlation of $\rho = 0.2$. For the autoregressive data, the model was of order one, and we again used two different correlations: a moderate correlation of $\rho = 0.7$ and a weak correlation of $\rho = 0.3$. For the unstructured data the variance–covariance matrix was as follows:

$$\mathbf{V} = \begin{bmatrix} 1.0 & 0.3 & 0.2 & 0.1 & 0.4 & 0.5 & 0.4 & 0.2 \\ 0.3 & 1.1 & 0.2 & 0.5 & 0.1 & 0.1 & 0.2 & 0.3 \\ 0.2 & 0.2 & 1.2 & 0.6 & 0.4 & 0.1 & 0.4 & 0.2 \\ 0.1 & 0.5 & 0.6 & 1.3 & 0.3 & 0.3 & 0.2 & 0.1 \\ 0.4 & 0.1 & 0.4 & 0.3 & 1.4 & 0.5 & 0.4 & 0.2 \\ 0.5 & 0.1 & 0.1 & 0.3 & 0.5 & 1.5 & 0.7 & 0.3 \\ 0.4 & 0.2 & 0.4 & 0.2 & 0.4 & 0.7 & 1.6 & 0.4 \\ 0.2 & 0.3 & 0.2 & 0.1 & 0.2 & 0.3 & 0.4 & 1.7 \end{bmatrix}. \tag{10}$$

This matrix corresponds to an outcome variable with an increasing variance (diagonal) and correlation between time points of between 0.07 $(= 0.1/\sqrt{1.3 \times 1.7})$ and 0.48 $(= 0.6/\sqrt{1.2 \times 1.3})$.

For the six data types we calculated the AIC, QIC_P, QIC_{HH} and DIC. For each criterion, the smallest value for the four different covariance structures was used to the select the "optimal" covariance. If the selected covariance was the known covariance, this was defined as a success.

Model fitting details

We used the SAS package to fit the mixed models and calculate the AIC, by using the MIXED procedure using restricted maximum likelihood and specifying the covariance structure using the REPEATED statement. The AIC was calculated using Equation (4) with p_A equal to the number of regression parameters plus the number of variance–covariance parameters.

We also used the SAS procedure GENMOD fit the GEE models, and calculated the QIC_P and QIC_{HH} using our own macro, which we verified by comparing to results in Hilbe (2009). The GENMOD procedure iteratively cycles between updating the regression parameters and updating the covariance parameters. The initial regression parameters are derived from a generalized linear model. However, the model often failed to converge when using an unstructured matrix. To overcome this problem, we altered the iterative procedure to update the covariance matrix once for every two updates of the regression parameters (using the RUPDATE=2 option in PROC GENMOD's REPEATED statement). All results were checked for convergence.

We used the WinBUGS package to fit the Bayesian models and calculate the DIC (Spiegelhalter et al, 2007). We used a burn-in of 3000 MCMC iterations followed by a sample of 3000 (Gelman et al, 2004, Chapter 11). To confirm the convergence of the MCMC samples we used the stationarity test of Heidelberger & Welch (1983). This test is available in the "coda" library of the R software package (Plummer et al, 2009). If the chain failed to converge, the model was re-run using the same data and the convergence re-checked.

We used vague priors for all unknown parameters. We used a vague prior for $\mathbf{V}$ by setting $\boldsymbol{\Sigma} = \mathbf{I}$ (the identity matrix), and $\nu = m$. We used a vague uniform prior for the autoregressive and exchangeable correlations: $\rho \sim \mathrm{U}(-1, 1)$. We used a uniform prior for the variance parameter for the autoregressive, exchangeable and independent correlations: $\sigma^2 \sim \mathrm{U}(0, 1000)$. For the autoregressive correlation the inverse-variance was calculated as $\tau = 1/\sigma^2$.

Empirical Data

We used data collected for a forest fragmentation study in the boreal forest of north-central Alberta, Canada (55° N, 113° W). Avian sampling was initiated in 1993, and conducted using 50- and 100-metre fixed-radius point-count plots in May and June of each year, over four to five visits per year. To account for species that used the plots but were not detected in some survey visits due to relatively low detectability (e.g., quiet or infrequent singing), we used total number of species observed over all rounds as the index of species richness (number of species observed per plot). In 1994, the study area was harvested to create three forest fragments in each 1 hectare, 10 hectare, 40 hectare and 100 hectare fragmentation treatment. An equal number and spatial distribution of sampling units in unharvested forest made up the controls for this experiment. Avian sampling was conducted annually through 2007, as the surrounding forest naturally regenerated. For additional sampling details see (Schmiegelow et al, 1997).

We used a subsample of the data for these analyses, representing 179 point count plots (clusters), each sampled annually for 15 years. Our total sample size was therefore 2865. We modeled effects of year, percent conifer within 200 meters of each point-count plot, and minimum June temperature, on avian species richness (number of avian species). A Q-Q plot was used to confirm that the response variable was approximately Normally distributed. Independent variables were selected for biological relevance, and to include time-variant, cluster-invariant, and cluster-variant variables. We used vague priors for all parameters in the Bayesian model, as in the simulation study. We used AIC, QIC_P, QIC_{HH} and DIC to compare the fit of the independent, exchangeable, autoregressive, and unstructured covariances, which described correlations among samples across years, within point-count plots.

Results

Simulation results

The percent successes from 100 simulations are shown in Table 1. The AIC performance was excellent when the true covariance structure was exchangeable or autoregressive (89%–100% correct). It had a high success rate for the independent covariance (70%–76% correct), but a low success rate for the unstructured covariance (13%–27% correct).

The QIC_{HH} gave almost identical results to the QIC_P, differing only by one for the unstructured matrix with a random covariate. Hin & Wang (2009) found similarly small differences when comparing these two versions of the QIC. From now on we refer to both statistics collectively as simply the "QIC". The QIC performed poorly when the true structure was independent or had a weak correlation (0%–14% correct). For these structures, the QIC most often incorrectly chose the unstructured covariance. This is the most complicated structure, as it uses the most covariance parameters. The QIC did much better for the moderately correlated autoregressive structure (81%–89% correct), but did poorly for the moderately correlated exchangeable (25%–30% correct), and only fairly well for the unstructured (40%–56% correct) covariances.

The DIC performance was excellent when the true covariance structure was exchangeable or autoregressive (92%–100% correct). It had a roughly 50% success for the independent (52%–58% correct) and unstructured (49%–50% correct) covariances. The convergence of the MCMC chains was generally very good, and less than 1% of the simulations needed to be re-fitted using more MCMC samples.

Combining the simulation results across the six data types and two covariate types, the overall success was 80.6% for the AIC, 29.4% for the QIC and 81.6% for the DIC.

To investigate further the performance of the methods we calculated the bias of the estimated regression and correlation parameters. The results are shown in Table 2. The average differences between the known and estimated parameters were small for every method and for both the correlation and regression parameters. This indicates that all three methods were equally unbiased at estimating the unknown parameters.

Empirical results

We focus on the statistical implications of our results, as the biological interpretation of more comprehensive models are addressed elsewhere (Schmiegelow et al., in prep). There are no strict rules about the significance of relative differences in AIC, QIC and DIC, but we can apply some guidelines. Burnham & Anderson (1998, page 70) consider a difference in the AIC of 10 to rule out the model with the larger AIC, and a difference of 0–2 to mean that the model fits are similar. Similarly, Hilbe (2009, page 260) considers a difference in the AIC of 0–2.5 to mean that the model fits are similar, and a difference greater than 10 to mean the model with the smaller AIC is preferred. These rules can equally be applied to the

QIC. A difference in the DIC of 5 is considered substantial, and a difference of 10 rules out the model with the larger DIC (Spiegelhalter et al, 2007).

Following these guidelines, the AIC and DIC both selected the unstructured covariance, which had the lowest value by more than 20 in both cases (Table 3). In contrast, the QIC indicated no difference between the independent, exchangeable and autoregressive structures, but ruled out the unstructured covariance as fitting the data poorly, as its QIC value was more than 10 units greater than QIC values for the other structures (Table 3).

The unstructured and exchangeable variance–covariance matrices estimated using the mixed model are shown in Fig 1. The x- and y-axes show the years 1993 to 2007 and the z-axis shows the covariances among responses at the same site but at different years. The covariances are always positive in this example. The ridge in the estimated variance–covariance along the diagonal represents the variance. The exchangeable correlation has a sharp fall from a variance of 9.4 to a constant covariance of 4.6 (hence the estimated within-cluster correlation is $4.6/9.4 = 0.49$). The estimated unstructured covariance is similar to but more variable than the exchangeable covariance, as it follows the basic pattern of a ridge and relatively little pattern with time lag among years.

To explore the unstructured covariance further, we plotted the average covariance (and 95% confidence intervals) by the distance between observations (in years) in Fig 2 (Grady, 1995). After a drop in the average covariance from observations in the same year to those 1 year apart, the covariance is reasonably stable to observations 7 years apart, and then declines. The correlation never declines to zero, even for the most distance observations.

Discussion

Simulation Study

Although we used three different models, they yielded equally unbiased estimates of the regression and correlation parameters (Table 2). Therefore the observed differences in the performance of the information criteria have little to do with the differences in estimation techniques, but are instead due to differences in the construction of the information criteria.

In our simulation study, the AIC and the DIC clearly outperformed the QIC in selecting the correct covariance structure (Table 1). The QIC did particularly badly when the true covariance structure was independent or had a weak exchangeable or autoregressive structure (0%–14% success). In these cases, the QIC was strongly biased towards selecting the unstructured covariance. This indicates that the QIC was not sufficiently penalizing the added complexity of the $m(m+1)/2$ parameters required for the unstructured covariance. To confirm this, we examined the trace from the QIC equation (5), as this part of the equation is designed to measure model complexity. Using the data with a weak autoregressive correlation as an example, most of the traces using an unstructured matrix were smaller than those using the three simpler matrices (independent, exchangeable and autoregressive). This explains why the QIC incorrectly ranked the complexity of the covariance structures. In contrast, the AIC (by design) and the DIC (by estimation) always correctly selected the largest number of parameters for the unstructured matrix. This gives the AIC and DIC an obvious advantage over the QIC.

For the autoregressive and exchangeable structures, the QIC did much better when there was a moderate correlation compared to a weak correlation. For the AIC and DIC there was only a small drop in performance when moving from a moderate to weak correlation (2%–11% drop for the AIC and 5%–6% for the DIC). The QIC needed a strong correlation in the data to work well, whereas the DIC worked well for both weak and moderate correlations. The AIC worked even better than DIC in most cases, except when the true covariance was unstructured. In that case, AIC was outperformed by both other criteria. This suggests that the AIC over-penalized the covariance parameters for the complex structures. As a result, the DIC might be preferable to the AIC when biological rationale cannot rule out the unstructured covariance, because it performed more consistently across a range of covariance structures. This difference occurred because the DIC uses the estimated number of parameters, whereas the AIC uses a fixed number of parameters (in this case 36 for an unstructured matrix). The Bayesian models often required fewer than 36 parameters to model the covariance matrix (10), which made an unstructured matrix more parsimonious and hence preferable.

The paper that introduced the QIC (Pan, 2001) contained a similar simulation study to that shown here. The study showed an approximate 70% success for the QIC in correctly selecting an exchangeable covariance (using $N = 50, 100$, $m = 3$ and $\rho = 0.5$). However, the study did not include the unstructured covariance as a possible alternative, and only used the independent, autoregressive and exchangeable structures. Also, the study did not look at correlations weaker than $\rho = 0.5$. Another simulation study found success rates for the QIC statistics of between 65% and 98%, but also did not include the unstructured covariance as a possible alternative (Hin et al, 2007). Based on the results of our study, the success rates for the QIC in these studies would have been lower if an unstructured covariance had been used, or if the data had been generated with a weaker correlation. Our results suggest that QIC is untrustworthy, and should not be used for selecting among competing covariance structures.

Empirical Data

The AIC and DIC both selected the unstructured covariance with the exchangeable correlation as second best, which appears reasonable based on the three-dimensional plot of the estimated unstructured covariance (Fig 1). The number of parameters used by the AIC and DIC agreed closely, while the number of parameters from the trace used by the QIC_P and QIC_{HH} were much smaller. As expected, the QIC_P and QIC_{HH} were the same for the independent and exchangeable models. In contrast, the fit of the AIC and DIC indicated a strong improvement in model fit between the independent and exchangeable models. Although the QIC statistics would lead us to conclude that there is no improvement in fit between the exchangeable and independent models, based on what we know about the data and territory selection in songbirds, this is implausible.

The QIC statistics tended to select overly complex structures in the simulation study. In contrast, both the QIC_P and QIC_{HH} selected the simpler autoregressive structure for the empirical data, whereas the AIC and DIC both indicated that the more complex unstructured covariance was best. An autoregressive structure creates a decay in correlation with increasing distance between years. This decay was estimated as $\rho = 0.51$. So observations of avian richness from the same location but one year apart are correlated by 0.51, and observations 2 years apart by $0.51^2 = 0.26$. Observations five years apart are only correlated by 0.03. This correlation structure therefore suggests that the similarity in avian richness is transitory and that neighboring years are the most important factor. In contrast, the unstructured correlation estimated that responses within 7 years of each other were roughly equally correlated, and that there was some decay in correlation thereafter (Fig 2). This implies that the persistent structural characteristics of each location are more likely to define its avian richness than richness in a previous year. This is biologically plausible, as many species are selective regarding forest structure, but show irruptive or highly temporally variable population sizes due to annual variation in reproductive success and overwintering mortality rates, which would be reflected in variable occupancy and resultant measures of avian species richness at the scale of individual plots.

The number of parameters used by the AIC and estimated number of parameters used by the DIC were almost identical (Table 3). The biggest difference was for the unstructured matrix where the DIC estimated 131.4 parameters and the AIC 137. The DIC used fewer parameters because of a positive correlation between the estimated parameters for this matrix, meaning that independent estimates were not needed. This is an advantage of the DIC over the AIC, as the DIC is able to estimate the actual complexity whereas the AIC relies on a fixed number of parameters. In this example the difference in complexity is small, and the ranking of the covariance matrices is the same for both criteria.

Given the considerations outlined above, we therefore concluded that the AIC and DIC were more likely to have selected a reasonable correlation structure, than the QIC.

Qualitative considerations

In addition to considering the relative performance of each approach, ecologists and practitioners need to consider which trade-offs, paradigms, and assumptions associated with each approach best meet their needs.

Generalized estimating equations are appealing for several reasons, including their relative simplicity (Fitzmaurice et al, 2004). Like generalized linear mixed models, they can accommodate any response distribution among the exponential family (Zorn, 2001). Further, both parameter estimates and empirical standard errors are robust to misspecification of the correlation structure (Overall & Tonidandel, 2004), the interpretation of the parameters is consistent when sample sizes vary (Pendergast et al, 1996), and GEEs are easily modeled using widely-available statistical packages (Fitzmaurice et al, 2004). They are therefore promising for ecological data that are clustered or longitudinal, but not Normally distributed. However, the QIC performed so poorly in our study that we cannot recommend this information criterion. Consequently GEEs should only be used when the biological rationale for selecting the covariance structure is obvious (see also a qualitative comparison that can be considered (Bishop et al, 2000)). We stress that our concerns are not with GEEs themselves, but of the problem of how to choose the best covariance when using GEEs. Ongoing work (mentioned below) is seeking to create a better criterion for identifying the covariance structure when using GEE models.

Other statistics

A number of other statistics (that we have not considered here) have been suggested to help select covariance structures. Hin & Wang (2009) proposed the correlation information criterion and also used the trace from Equation (7), and found that both were substantially better at selecting the correct covariance compared with the QIC. Hilbe (2009, Section 13.2.4) gave a useful discussion on the AIC, QIC and trace, and provided some empirical evidence for the trace statistic outperforming the QIC. Shults et al (2009) compared the Rotnitzky-Jewell, DBAR, simple and rule-out criterion. They found that the Rotnitzky-Jewell statistic performed best at identifying an autoregressive structure. Lastly, we know that Hilbe is currently working on developing a more accurate information criterion (Hilbe 2009, personal communication).

Limitations of this study

The study compared three different criteria associated with different statistical modelling approaches. The AIC is based on a classical statistical approach and maximum likelihood, while the QIC is also based on a classical statistical approach but with the quasi-likelihood. The DIC results from a Bayesian approach and MCMC inference. Despite the different methods, the goal for all three criteria is the same: to identify the best covariance structure. This is often of practical interest to researchers. Hence we feel it is important that they are aware of the limitations and benefits of the QIC, AIC and DIC.

In our simulation study we did not consider the size of difference between the best criterion value and the next best, but simply chose the covariance structure associated with the smallest criterion value. In practice if two different covariance structures have similar criterion values then it could be misleading to assume the covariance with the smallest value gives the best fit. When this happens it is best to report the results of both models, or, if the inferences are similar, the most parsimonious model.

We used the fixed effects AIC (4) which over-estimated the number of parameters when using an unstructured covariance in our simulation study. There is an adjusted version of the AIC to compensate for random effects (Vaida & Blanchard, 2005), but not to compensate for correlated parameters in the variance–covariance matrix of the residuals. A version of the AIC that incorporated this adjustment would likely perform better at selecting the correct variance–covariance structure. Despite this flaw the AIC still performed much better than the QIC in our simulation study.

Summary and recommendations

Our study compared three different methods for selecting the correct covariance structure for ecological modeling. The results showed that the DIC was a better all-round statistic for making this choice, although it was out-performed by the AIC when the true structure was independent. The overall success rates of the AIC and DIC were similar. However, we recommend using the DIC because it adjusts for correlated parameters when using the unstructured variance–covariance, whereas the version of the AIC used here does not. When using the AIC to compare models with missing covariate data, it would be preferable to adjust for any changes in sample size by using the modified version of the AIC discussed in Hilbe (2009, Section 7.3).

We cannot recommend the use of the QIC, as our simulation study showed it did not sufficiently penalize complex covariances, and so often wrongly selected more complex models.

Acknowledgements

Computational resources and services used in this work were provided by the High Performance Computer and Research Support Unit, Queensland University of Technology, Brisbane, Australia. We thank L. Lix and W. Pan for statistical advice, and K. Aitken for data support.

References

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723.

Bishop J, Die D, Wang YG (2000) A generalized estimating equations approach for analysis of the impact of new technology on a trawl fishery. *Australian and New Zealand Journal of Statistics*, 42, 159–177.

Burnham KP, Anderson DR (1998) *Model Selection and Inference; A Practical Information-Theoretic Approach*. Springer-Verlag, New York, USA.

Clifford P, Richardson S, Hémon D (1989) Assessing the significance of the correlation between two spatial processes. *Biometrics*, 45, 123–134.

Diggle PJ, Heagerty P, Liang KY, Zeger SL (2002) *Analysis of Longitudinal Data*, 2nd edn. Oxford University Press, Oxford.

Dobson AJ, Barnett AG (2008) *An Introduction to Generalized Linear Models*, 3rd edn. Chapman & Hall/CRC, Boca Raton, Fla.

Dreitz VJ, Kitchens WM, DeAngelis DL (2004) Effects of natal departure and water level on survival of juvenile snail kites (Rostrhamus sociabilis) in Florida. *Auk*, 121, 894–903.

Driscoll MJL, Donovan T, Mickey R, Howard A, Fleming KK (2005)

Determinants of wood thrush nest success: a multi-scale, model selection approach. *Journal of Wildlife Management*, 69, 699–709.

Duncan RP (2004) Extinction and endemism in the New Zealand avifauna. *Global Ecology and Biogeography*, 13, 509–517.

Fitzmaurice GM, Laird NM, Ware JH (2004) *Applied Longitudinal Analysis*. John Wiley & Sons, Hoboken, New Jersey, USA.

Gelman A, Carlin JB, Stern HS, Rubin DB (2004) *Bayesian Data Analysis*, 2nd edn. Chapman & Hall/CRC, Boca Raton, Fla.

Gillies CS, Hebblewhite M, Nielsen SE, Krawchuk MA, Aldridge CL, Frair JL, Saher DJ, Stevens CE, Jerde CL (2006) Application of random effects to the study of resource selection by animals. *Journal of Animal Ecology*, 75, 887–898.

Grady JJ, Helms RW (1995) Model selection techniques for the covariance matrix for incomplete longitudinal data. *Statistics in Medicine*, 14, 1397–1416.

Hardin JW, Hilbe JM (2003) *Generalized Estimating Equations*. Chapman and Hall, New York.

Heidelberger P, Welch PD (1983) Simulation run length control in the presence of an initial transient. *Opns Res*, 31, 1109–44.

Helser TE, Stewart IJ, Lai HL (2007) A Bayesian hierarchical meta-analysis of growth for the genus Sebastes in the eastern Pacific ocean. *Canadian Journal of Fisheries and Aquatic Science*, 64, 470–485.

Hilbe, JM (2009) *Logistic Regression Models*, Chapman & Hall/CRC Press, Boca Raton, FL.

Hin L-Y, Carey VJ, Wang Y-G (2007) Criteria for Working-Correlation-Structure Selection in GEE: Assessment via Simulation. *The American Statistician*, 61, 360–364.

Hin L-Y, Wang Y-G (2009) Working-correlation-structure identification in generalized estimating equations. *Statistics in Medicine*, 28, 642–658.

Koper N, Schmiegelow FKA (2006) Effects of habitat management for ducks on target and non-target species. *Journal of Wildlife Management*, 70, 823–834.

Krawchuk MA, Taylor PD (2003) Changing importance of habitat structure across multiple spatial scales for three species of insects. *Oikos*, 103, 153–161.

Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22.

Overall JE, Tonidandel S (2004) Robustness of generalized estimating equation (GEE) tests of significance against misspecification of the error structure model. *Biometrical Journal*, 46, 203–213.

Pan W (2001) Akaike's information criterion in generalized estimating equations. *Biometrics*, 57, 120–125.

Pendergast JF, Gange SJ, Newton MA, Lindstrom MJ, Palta M, Fisher MR (1996) A survey of methods of analyzing clustered binary response data. *International Statistical Review*, 64, 89–118.

Plummer M, Best N, Cowles K, Vines K (2009) coda: Output analysis and diagnostics for MCMC. R package version 0.13-4.

Reynolds (2004) Alterable Predictors of Child Well-Being in the Chicago Longitudinal Study. *Children and Youth Services Review*, 26, 1–14.

Schmiegelow FKA, Machtans CS, Hannon SJ (1997) Are boreal birds resilient to forest fragmentation? An experimental study of short-term community responses. *Ecology*, 78, 1914–1932.

Schneider SK, Law R, Illian JB (2006) Quantification of neighbourhood-dependent plant growth by Bayesian hierarchical modelling. *Journal of Ecology*, 94, 310–321.

Shults J, Sun W, Tu X, Kim H, Amsterdam J, Hilbe JM, Ten-Have T (2009) A comparison of several approaches for choosing between working correlation structures in generalized estimating equation analysis of longitudinal binary data. *Statistics in Medicine*, 28, 2338-2355.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B*, 64, 583–640.

Spiegelhalter DJ, Thomas A, Best NG, Lunn D (2007) WinBUGS version 1.4.2 user manual.

Vaida F, Blanchard S (2005) Conditional Akaike information for mixed-effects models. *Biometrika*, 92, 351–370.

Wiens JA (1989) Spatial scaling in ecology. *Functional Ecology*, 3, 385–397.

Zorn CJW (2001) Generalized estimating equation models for correlated data: a review with applications. *American Journal of Political Science*, 45, 470–490.

Table 1: Comparison of the (a) AIC, (b) QIC_P, (c) QIC_{HH} and (d) DIC for selecting the true covariance structure for a model with either a fixed or random covariate, using simulated data. Cells show the percent of successful selections. Numbers in bold show the percent of correct choices.

(a) Results for the AIC

True covariance	Fixed covariate: $X_{it} = t$ — Selected covariance: Indep.	Exch.	AR	Unst.	Random covariate: $X_{it} \sim \mathrm{N}(0,1)$ — Selected covariance: Indep.	Exch.	AR	Unst.
Independent	**70**	15	15	0	**76**	14	9	1
Exchangeable ($\rho = 0.2$)	0	**97**	2	1	0	**98**	2	0
Exchangeable ($\rho = 0.5$)	0	**100**	0	0	0	**100**	0	0
Autoregressive ($\rho = 0.3$)	0	3	**97**	0	1	10	**89**	0
Autoregressive ($\rho = 0.7$)	0	0	**100**	0	0	0	**100**	0
Unstructured	0	49	24	**27**	0	60	27	**13**

(b) Results for the QIC_P

True covariance	Fixed covariate: $X_{it} = t$ — Selected covariance: Indep./Exch.†	AR	Unst.	Random covariate: $X_{it} \sim \mathrm{N}(0,1)$ — Selected covariance: Indep.	Exch.	AR	Unst.
Independent	**3**	0	97	**2**	4	5	89
Exchangeable ($\rho = 0.2$)	**3**	3	94	0	**0**	0	100
Exchangeable ($\rho = 0.5$)	**25**	7	68	0	**30**	13	57
Autoregressive ($\rho = 0.3$)	3	**14**	83	0	2	**10**	88
Autoregressive ($\rho = 0.7$)	7	**81**	12	0	4	**89**	7
Unstructured	17	27	**56**	5	22	33	**40**

(c) Results for the QIC_{HH}

True covariance	Fixed covariate: $X_{it} = t$ — Selected covariance: Indep./Exch.†	AR	Unst.	Random covariate: $X_{it} \sim \mathrm{N}(0,1)$ — Selected covariance: Indep.	Exch.	AR	Unst.
Independent	**3**	0	97	**2**	4	5	89
Exchangeable ($\rho = 0.2$)	**3**	3	94	0	**0**	0	100
Exchangeable ($\rho = 0.5$)	**25**	7	68	0	**30**	13	57
Autoregressive ($\rho = 0.3$)	3	**14**	83	0	2	**10**	88
Autoregressive ($\rho = 0.7$)	7	**81**	12	0	4	**89**	7
Unstructured	17	27	**56**	5	21	34	**40**

(d) Results for the DIC

True covariance	Fixed covariate: $X_{it} = t$ — Selected covariance: Indep.	Exch.	AR	Unst.	Random covariate: $X_{it} \sim \mathrm{N}(0,1)$ — Selected covariance: Indep.	Exch.	AR	Unst.
Independent	**58**	30	11	1	**52**	25	22	1
Exchangeable ($\rho = 0.2$)	0	**95**	1	4	0	**94**	4	2
Exchangeable ($\rho = 0.5$)	0	**100**	0	0	0	**99**	0	1
Autoregressive ($\rho = 0.3$)	0	5	**92**	3	1	2	**93**	4
Autoregressive ($\rho = 0.7$)	0	0	**98**	2	0	0	**99**	1
Unstructured	0	39	12	**49**	0	32	18	**50**

† The QIC_P and QIC_{HH} both give identical results for an independent and exchangeable covariance when using the sandwich covariance matrix without a subject-specific and time-independent covariate

Table 2: Bias in the regression correlation parameter estimates for the four information criteria using simulated data. Results for a fixed covariate $X_{it} = t$ with a known regression parameter of $\beta = 0.3$. Cells show the mean and standard deviation of the bias for the 100 simulations.

(a) Results for the AIC

	Regression parameter	Correlation parameter
Independent	0.006 (0.028)	NA
Exchangeable ($\rho = 0.2$)	−0.002 (0.029)	0.003 (0.065)
Exchangeable ($\rho = 0.5$)	0.001 (0.028)	−0.011 (0.088)
Autoregressive ($\rho = 0.3$)	−0.005 (0.034)	−0.001 (0.061)
Autoregressive ($\rho = 0.7$)	−0.002 (0.031)	−0.005 (0.048)

(b) Results for the QIC_P and QIC_{HH}

	Regression parameter	Correlation parameter
Independent	0.006 (0.028)	NA
Exchangeable ($\rho = 0.2$)	−0.002 (0.029)	−0.005 (0.064)
Exchangeable ($\rho = 0.5$)	0.001 (0.028)	−0.022 (0.087)
Autoregressive ($\rho = 0.3$)	−0.005 (0.034)	−0.007 (0.062)
Autoregressive ($\rho = 0.7$)	−0.002 (0.031)	−0.009 (0.053)

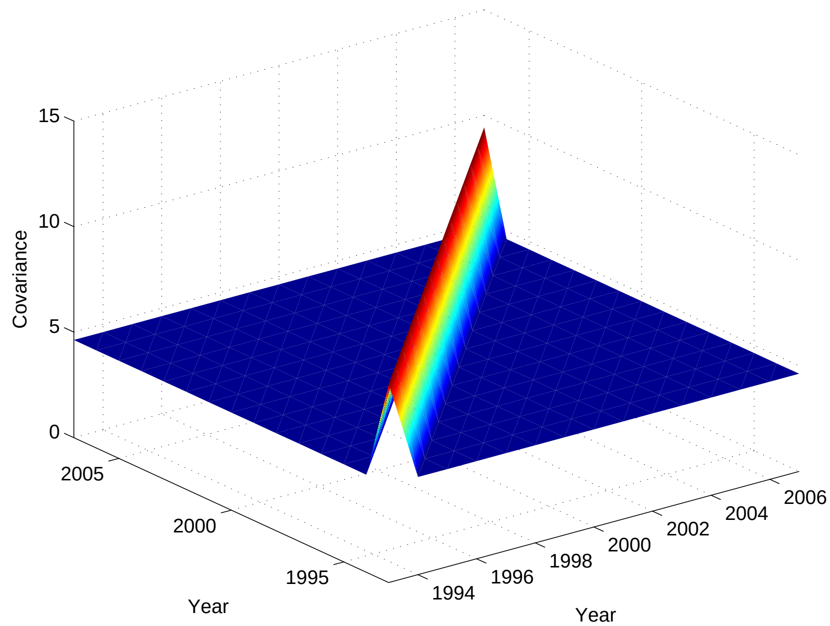
(c) Results for the DIC

	Regression parameter	Correlation parameter
Independent	0.005 (0.028)	NA
Exchangeable ($\rho = 0.2$)	−0.002(0.029)	−0.014 (0.134)
Exchangeable ($\rho = 0.5$)	0.002 (0.029)	−0.014 (0.091)
Autoregressive ($\rho = 0.3$)	−0.005 (0.033)	0.001 (0.069)
Autoregressive ($\rho = 0.7$)	−0.002 (0.031)	−0.015 (0.084)

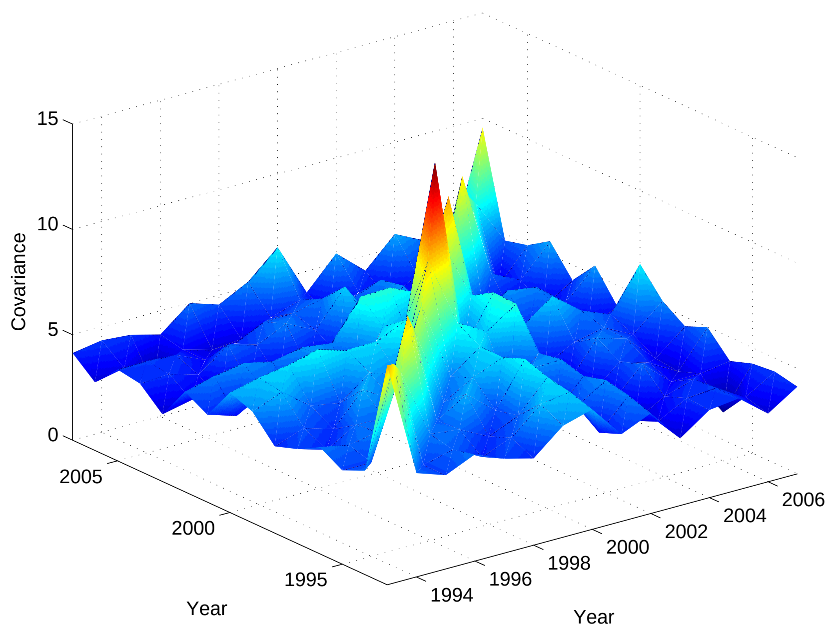
Table 3: Using the AIC, QIC_P, QIC_{HH} and DIC for choosing the optimal covariance structure for modeling long-term data from a forest fragmentation study (Schmiegelow et al, 1997). Smaller values of the criteria indicate a better fit. Best values for each criteria highlighted in bold font.

	Independent	Exchangeable	Autoregressive	Unstructured
(a) AIC				
-2 Log L	13603	12299	12822	12041
No. of parameters†	18	19	19	137
AIC values	13639	12337	12860	**12315**
(b) QIC_P				
$-2Q(\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}, \mathbf{I})$	2668.1	2668.1	2667.9	2660.6
Trace	28.5	28.5	28.5	38.1
QIC_P values	2725.1	2725.1	**2724.8**	2736.9
(c) QIC_{HH}				
$-2Q(\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}, \mathbf{I})$	2668.1	2668.1	2667.9	2660.6
Trace	28.5	28.5	28.5	38.3
QIC_{HH} values	2725.1	2725.1	**2724.8**	2737.3
(d) DIC				
$D(\mathbf{Y}\|\overline{\boldsymbol{\beta}})$	13614	12302	12832	12057
Estimated no. of parms. (pD)†	18.0	19.2	19.0	131.4
DIC values	13650	12340	12870	**12320**

† Number of parameters used by the regression model and variance–covariance matrix, estimated number of parameters for the DIC



(A)



(B)

Figure 1: Three-dimensional plots of the: (A) estimated exchangeable, and (B) unstructured variance-covariance matrices estimated using the mixed model for modeling long-term data from a forest fragmentation study (Schmiegelow et al, 1997).

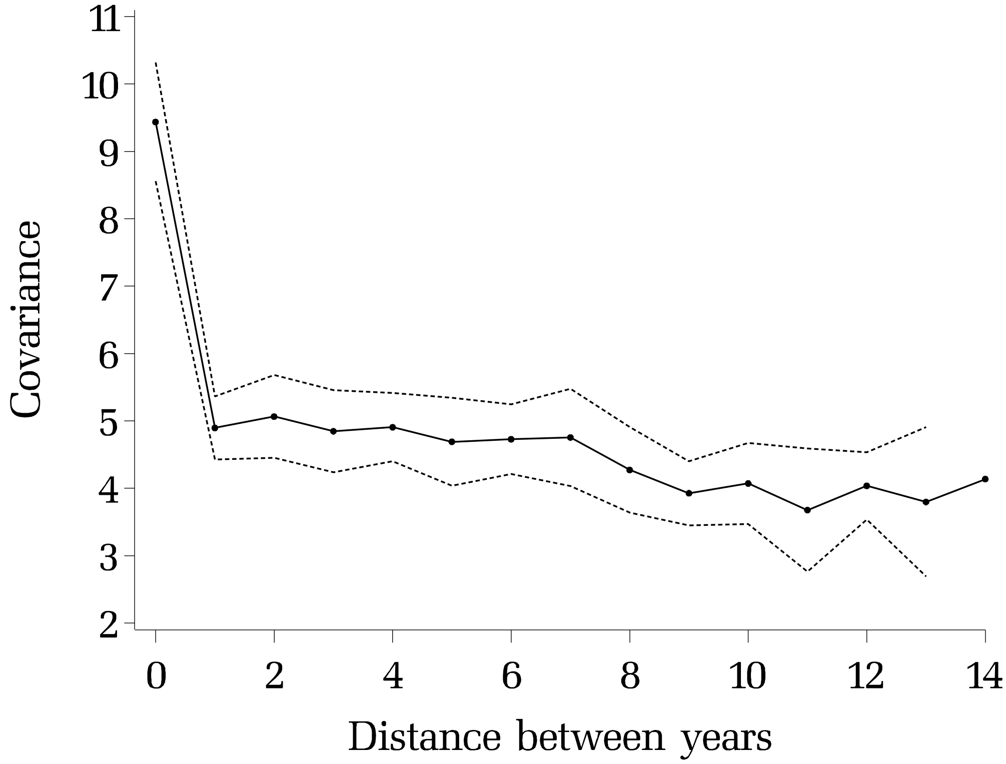


Figure 2: Average covariances (and 95% confidence interval) by distance between years for the unstructured variance-covariance matrix from Figure 1. Estimates from the mixed model.